

# **Linear Models in Medical Imaging**

**John Kornak**

**MI square**

**February 23, 2010**

# **Acknowledgement / Disclaimer**

**Many of the slides in this lecture have been adapted from slides available in talks available on the SPM web site.**

# Overview

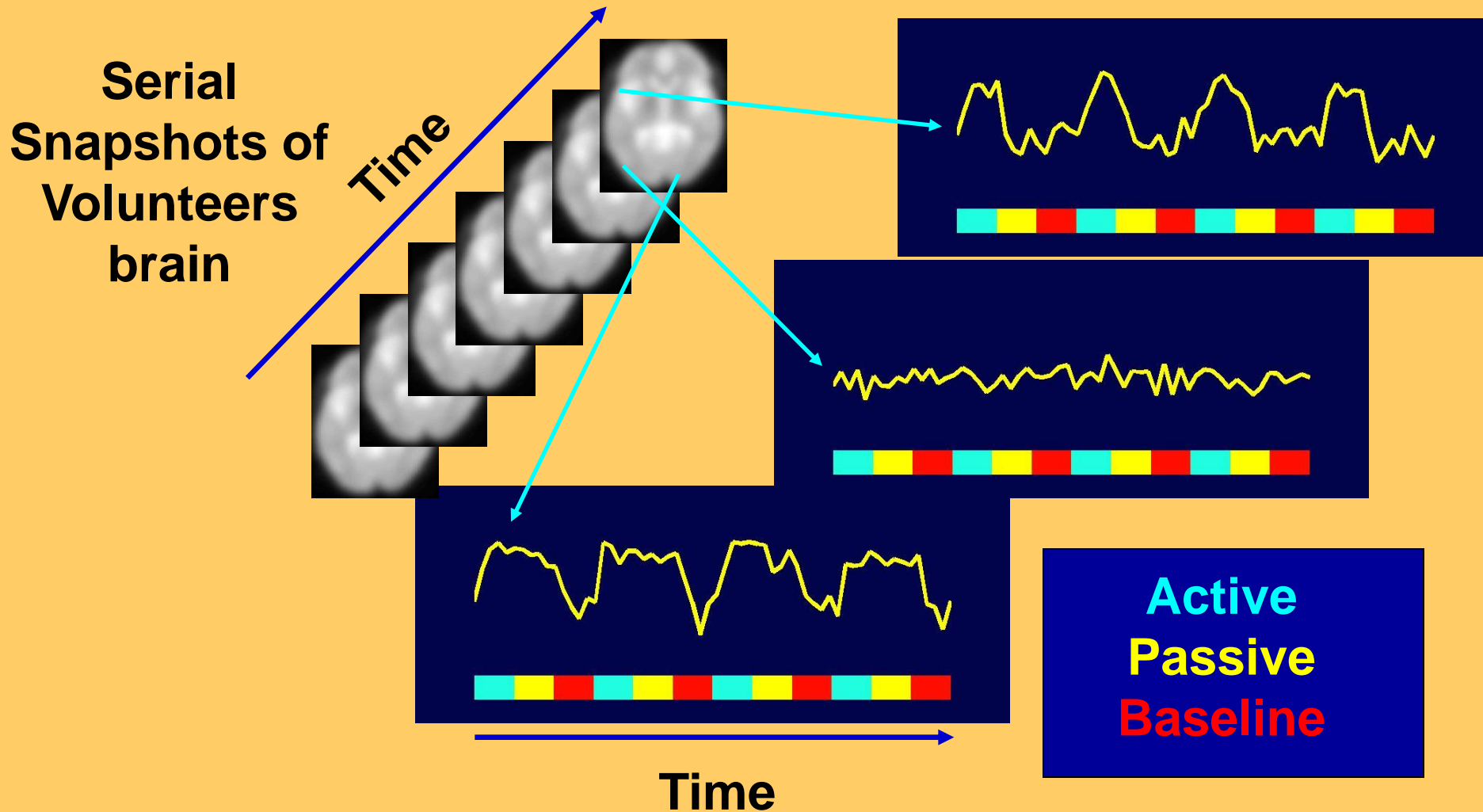
- **Motivation**
- **Linear model formulation**
- **Region of interest analyses**
- **Pixel/voxel based analyses**
- **Multiple comparisons for images**
- **Bayesian image analysis methods**

# Motivation

- **Imaging data – statistical methods to look for “regional effects”**
- **Tissue differences between groups or over time – VBM, TBM (voxel/tensor-based morphometry)**
- **PET (positron emission tomography), fMRI (functional MRI) – determine “activation” in the brain due to thought, stimulus or task**
- **Diffusion (DWI, DTI, tractography), Bone mineral density etc. etc.**

# FMRI Data:

**Set of Volumes (over time) or  
Set of Time-Series (over space)**



# Software etc.

**SPM – PET, fMRI, VBM and TBM, EEG/MEG**  
**(<http://www.fil.ion.ucl.ac.uk/spm/> needs Matlab)**

**FSL – fMRI primarily + DTI**  
**(<http://www.fmrib.ox.ac.uk/fsl/>)**

**R – Analyze fMRI package + linear models in general**  
**(<http://www.r-project.org/> and then go to your nearest CRAN mirror)**

**Also, check “Venables and Ripley” Splus book + many R books (see R web site) + online tutorials**

# Challenges

- **Generating suitable (statistical) imaging models**
- **Dealing with highly multivariate responses (curse of dimensionality)**
- **Defining imaging “hypotheses”**
- **Creating computationally efficient analysis procedures**

# **Aims of Statistical Modeling**

- **Summarize data**
- **Estimation: point and interval estimates**
- **Inference: hypotheses / relationships**
- **Prediction**

# Aims of Statistical Modeling

- **Summarize data**
- **Estimation: point and interval estimates**
- **Inference: hypotheses / relationships**
- **Prediction**

# Statistical Modeling Strategy

- **Propose a model for the data**
- **Fit the model**
- **Assess the model's adequacy**
- **Fit other plausible models**
- **Compare all fitted models**
- **Interpret the best model**

# Statistical Models: Definitions

- **Univariate response variable**  $y_i$  (for exp. unit  $i$ )
- **Covariates**  $(x_{i1}, x_{i2}, \dots, x_{ik}) = \mathbf{X}_i^T$   
(variables of interest and “nuisance” variables)
- **Data is:**  $\{y_i, \mathbf{x}_i^T; i = 1, \dots, n\}$ ,  $n$  experimental units

***Continuous covariates:*** e.g. age, blood pressure etc., (random or controlled)

***Factors:*** e.g. diagnosis, gender, drinking level (low, medium, high) etc.

# The (General) Linear Model

A simple *linear model* might take the form:

$$y_i = \beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + \dots + x_{im}\beta_m + \varepsilon_i$$

**e.g.**

$$y_i = \beta_{mean} + x_{i,age}\beta_{age} + x_{i,gender}\beta_{gender} + \dots + x_{i,diagnosis}\beta_{diagnosis} + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2), \quad i.i.d. \quad i = 1, \dots, n$$

*i.i.d.* = independently and identically distributed

# The (General) Linear Model

For univariate data:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n$$

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^T$$

is a set of unknown parameters

or in matrix notation

$$\mathbf{y} = \mathbf{X}^T \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

This can be extended to a multivariate response

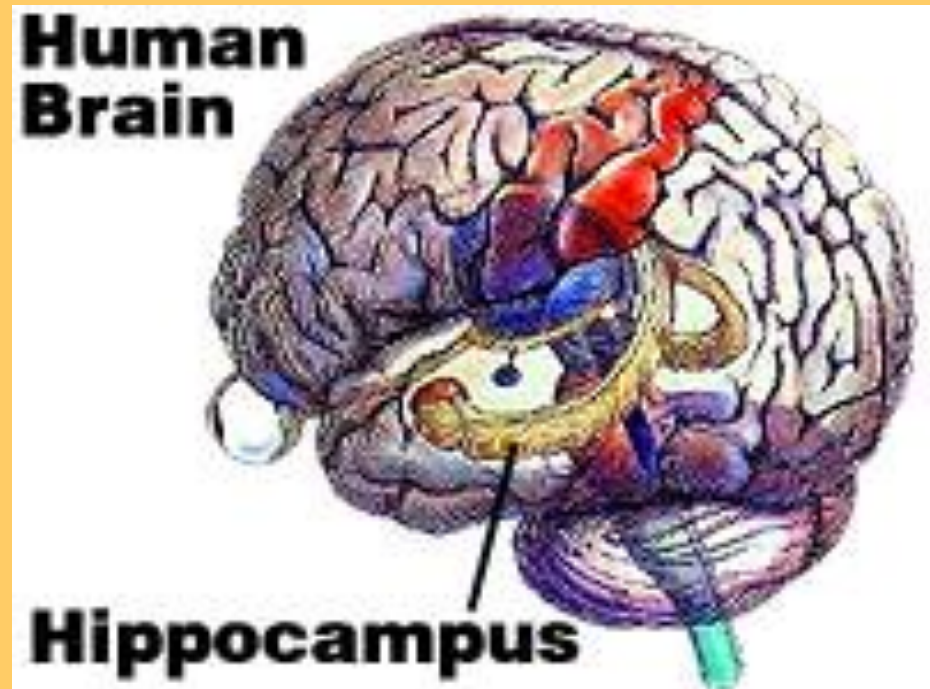
$$\mathbf{Y} = \mathbf{X}^T \mathbf{B} + \mathbf{E}$$

# Ex. Hippocampal Volume

HCV ~ Age + Diagnosis

(Wilkinson notation)

Diagnosis can be  
normal control  
(NC) or  
Alzheimer's  
disease (AD)

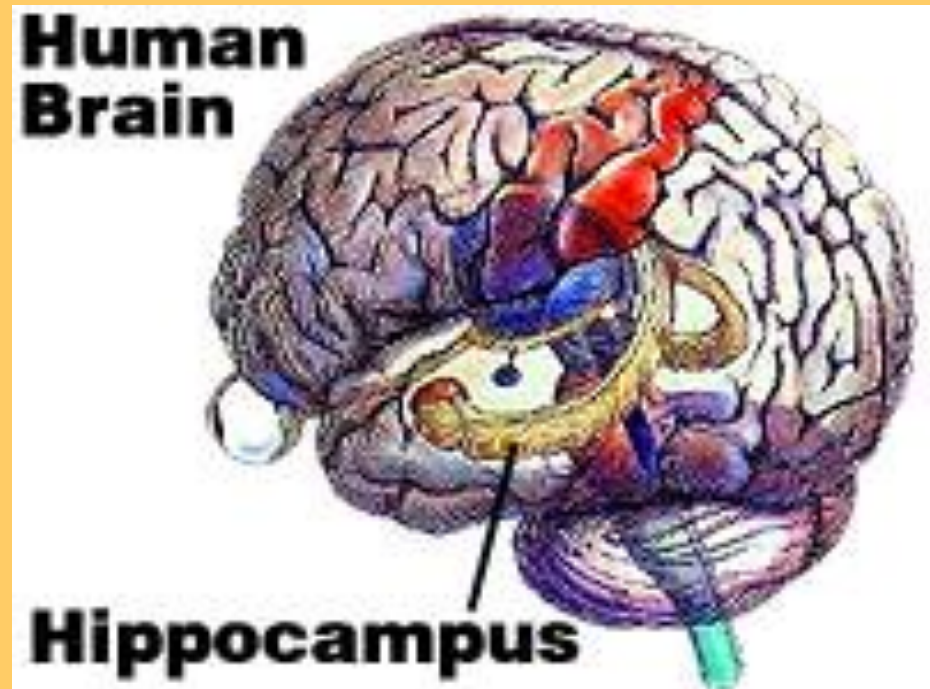


# Ex. Hippocampal Volume

$HCV \sim \text{Age} + \text{Diagnosis} + \text{Age} * \text{Diagnosis}$

(Wilkinson notation)

Diagnosis can be  
normal control  
(NC) or  
Alzheimer's  
disease (AD)



**Structural T1 weighted MRI's**

**Hippocampal volumes manually traced**

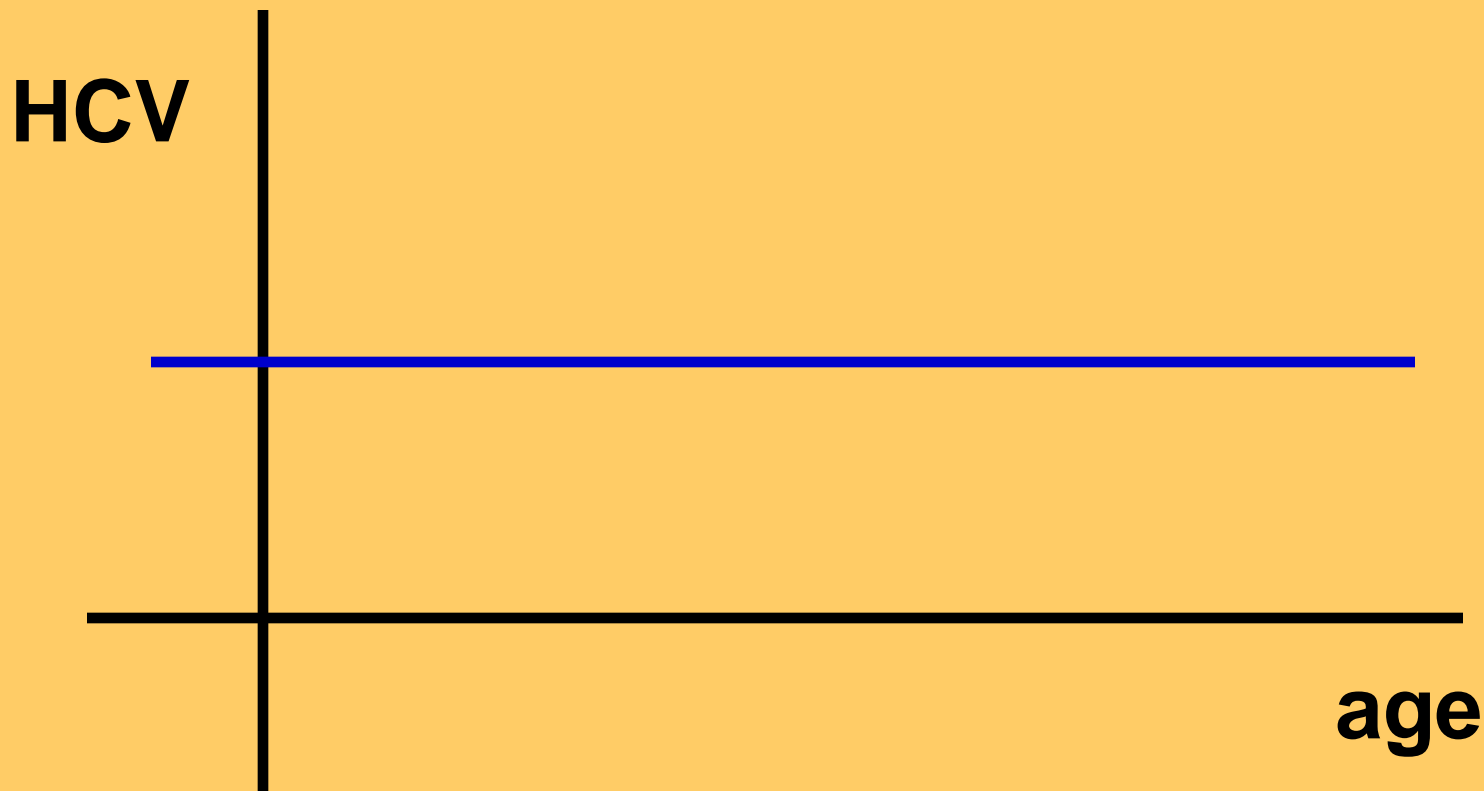
**Volume measure = response for each subject**

**Disease status encoded 1 for AD and 0 for NC  
(the  $x_{diag.}$  term)**

$$y_i = \beta_1 + x_{i,age}\beta_{age} + x_{i,diag.}\beta_{diag.} + x_{i,age}x_{i,diag.}\beta_{inter} + \varepsilon_i$$

**Case 1**

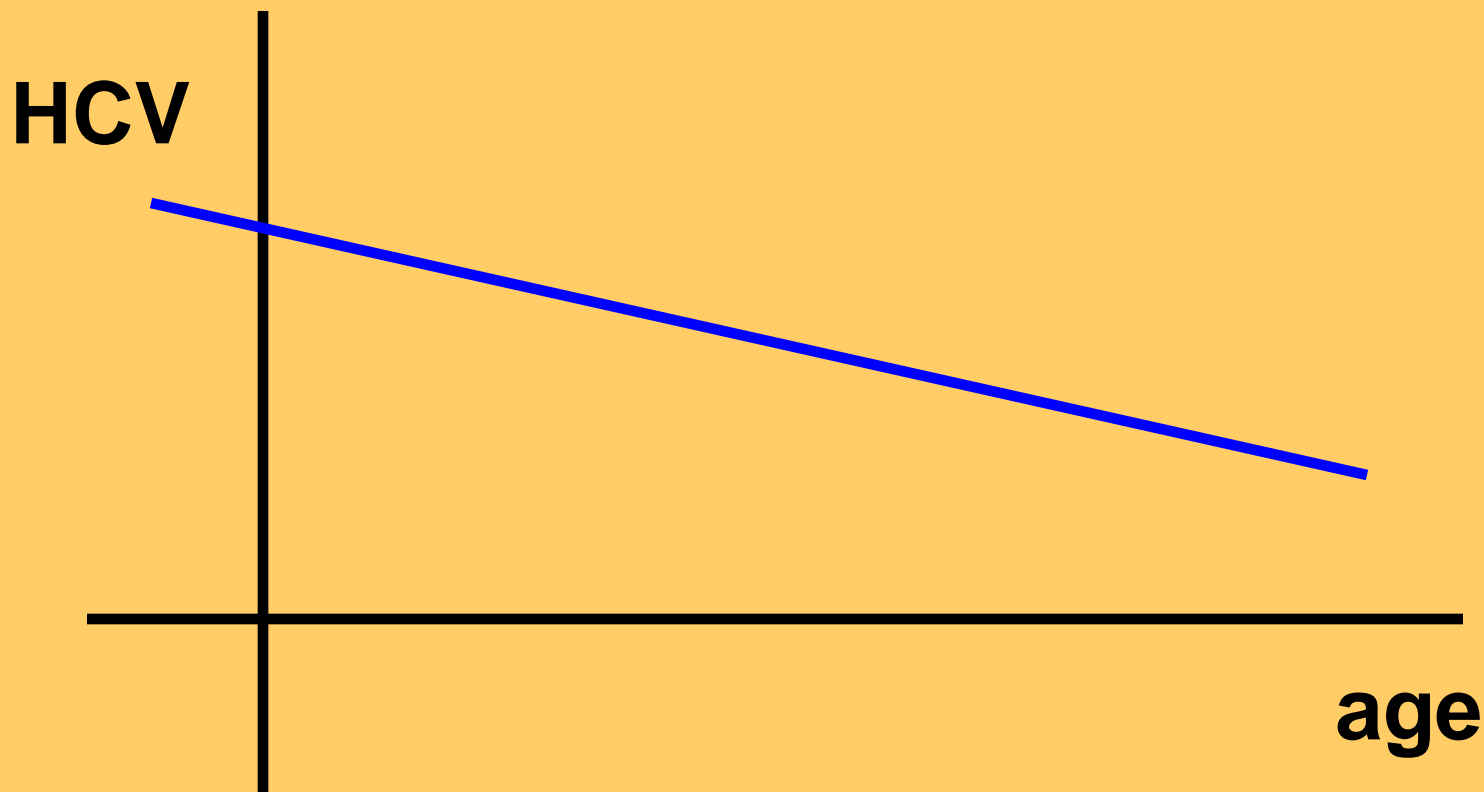
$$\beta_{age} = 0, \beta_{diag.} = 0, \beta_{inter} = 0$$



$$y_i = \beta_1 + x_{i,age}\beta_{age} + x_{i,diag.}\beta_{diag.} + x_{i,age}x_{i,diag.}\beta_{inter} + \varepsilon_i$$

**Case 2**

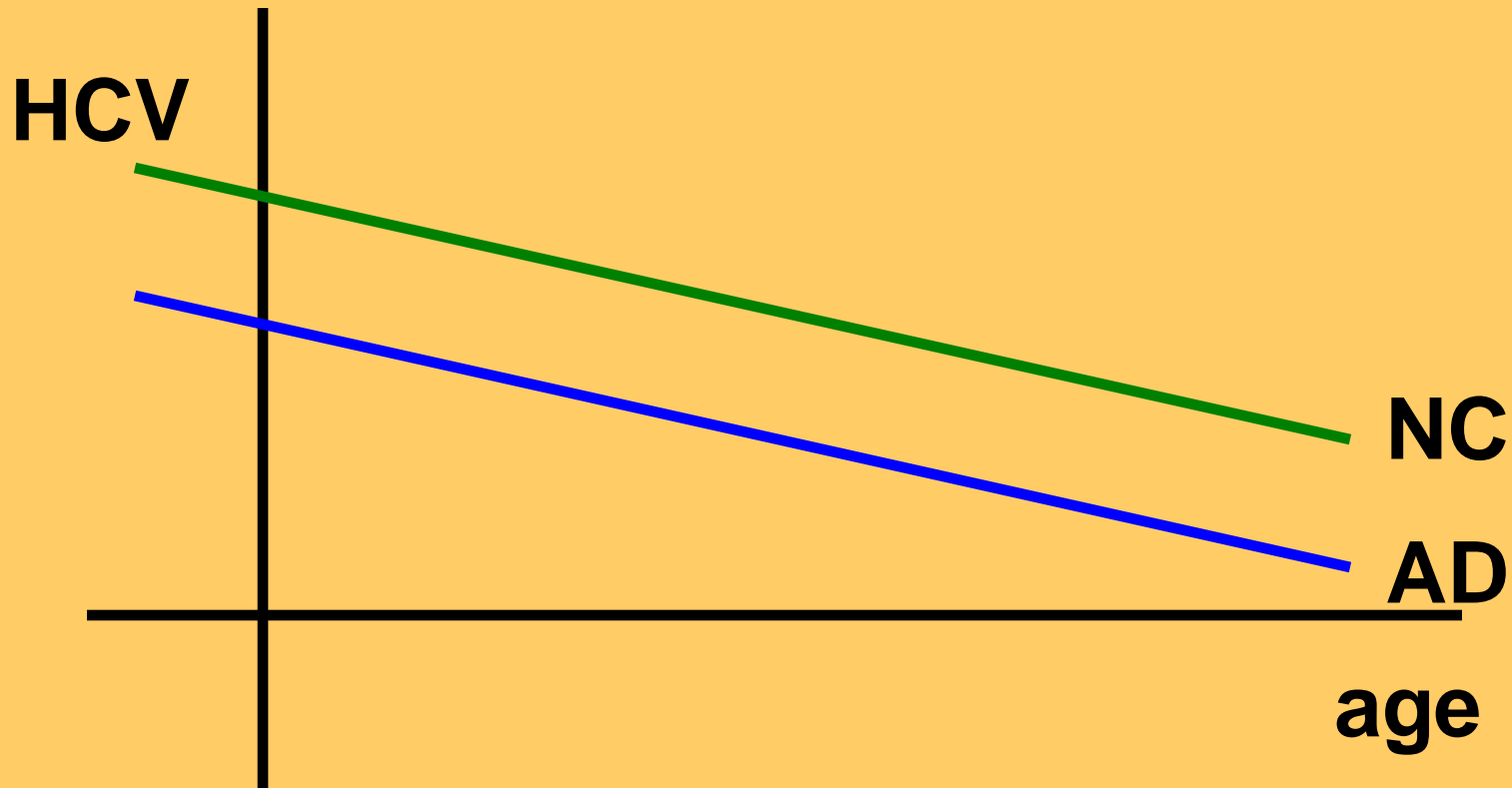
$$\beta_{age} \neq 0, \beta_{diag.} = 0, \beta_{inter} = 0$$



$$y_i = \beta_0 + x_{i,age}\beta_{age} + x_{i,diag.}\beta_{diag.} + x_{i,age}x_{i,diag.}\beta_{inter} + \varepsilon_i$$

### Case 3

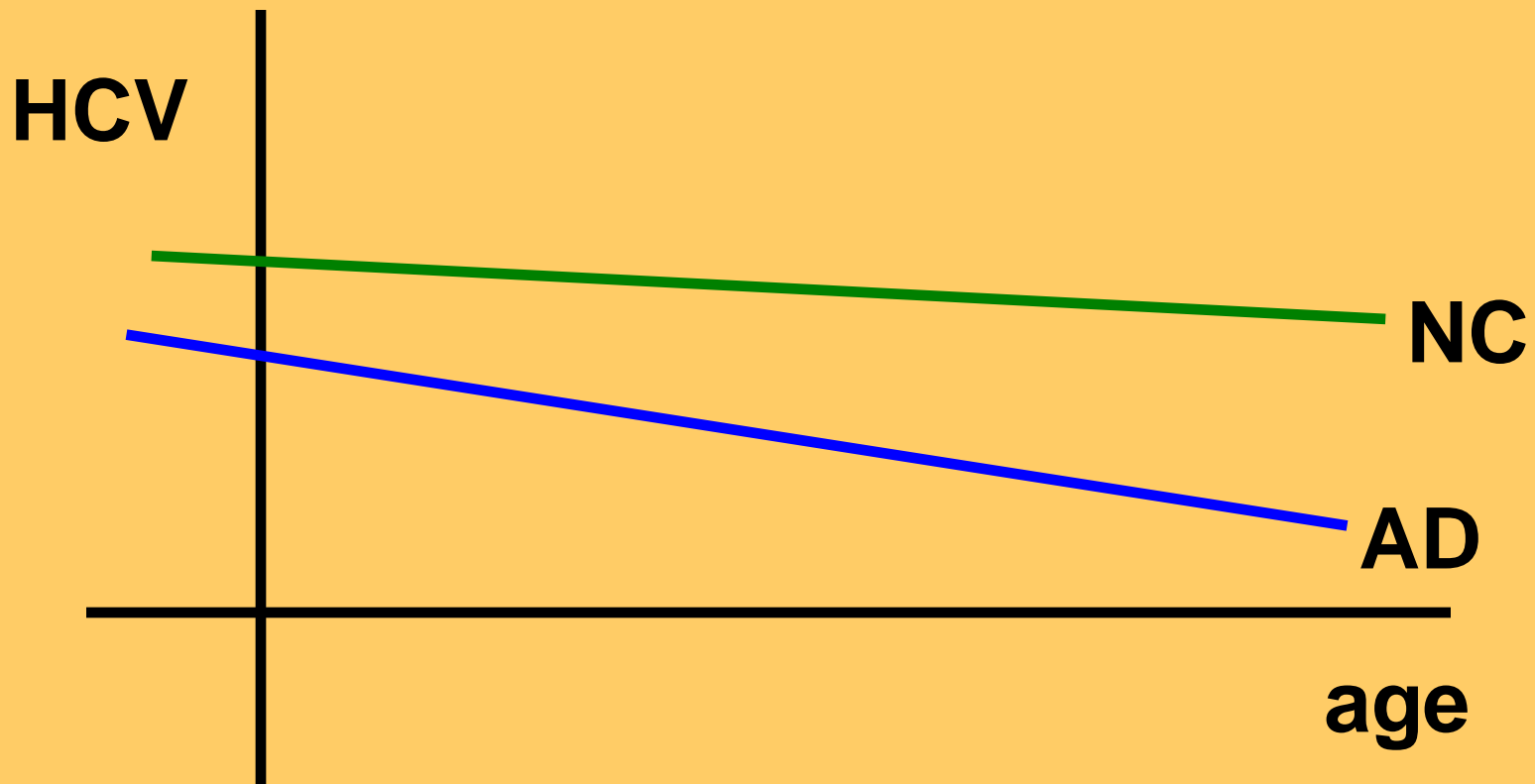
$$\beta_{age} \neq 0, \beta_{diag.} \neq 0, \beta_{inter} = 0$$



$$y_i = \beta_0 + x_{i,age}\beta_{age} + x_{i,diag.}\beta_{diag.} + x_{i,age}x_{i,diag.}\beta_{inter} + \varepsilon_i$$

**Case 4**

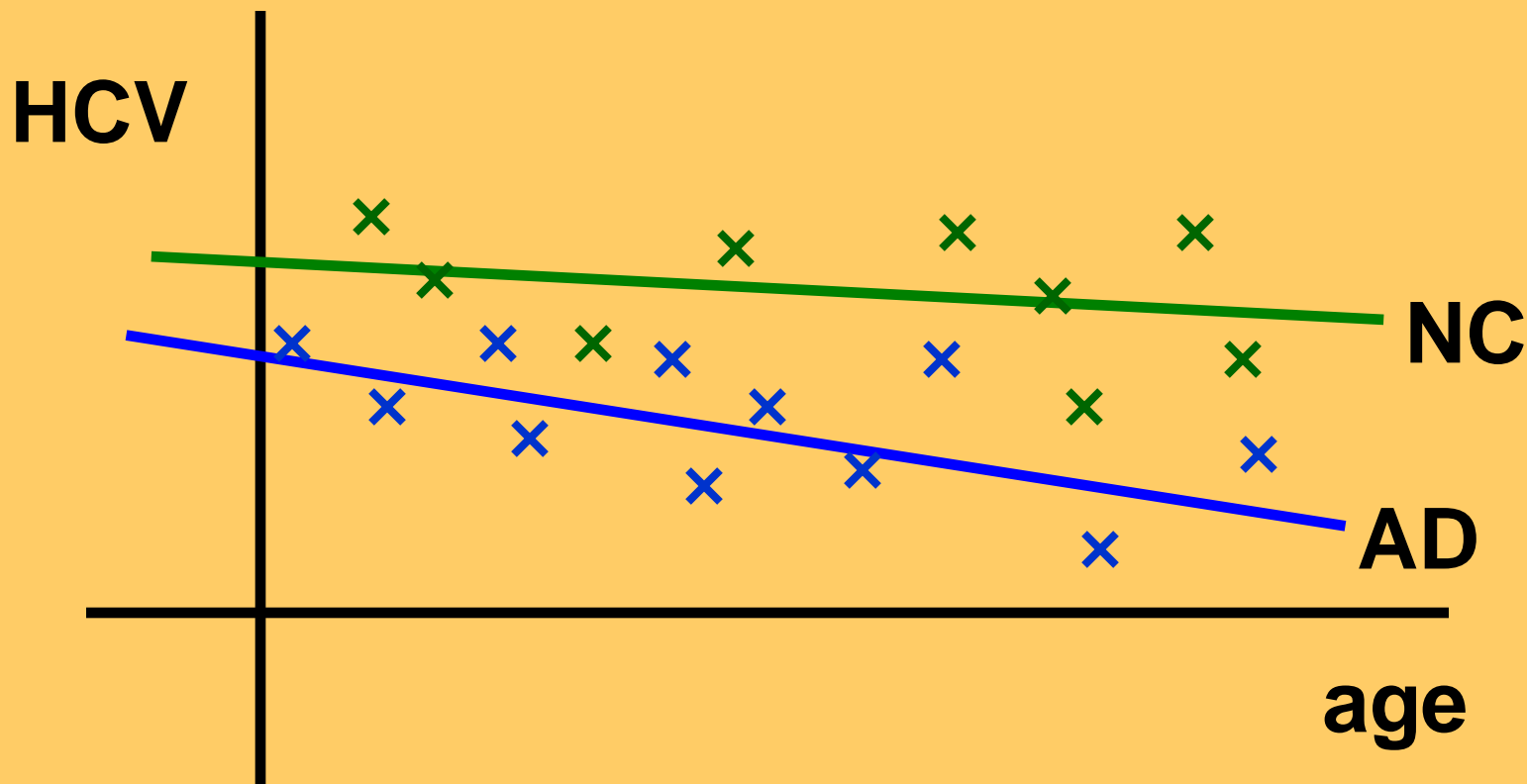
$$\beta_{age} \neq 0, \beta_{diag.} \neq 0, \beta_{inter} \neq 0$$



$$y_i = \beta_1 + x_{i,age}\beta_{age} + x_{i,diag.}\beta_{diag.} + x_{i,age}x_{i,diag.}\beta_{inter} + \varepsilon_i$$

**Case 4**

$$\beta_{age} \neq 0, \beta_{diag.} \neq 0, \beta_{inter} \neq 0$$



**Linear models can be more general**

**- only needs to be linear in the parameters:  $\beta$**

**We can have:**

$$y_i = x_{age}\beta_1 + x_{age}^2\beta_2 + \exp(x_{height})\beta_3 + x_{age}^\pi x_{height}\beta_4 + \varepsilon_i$$

$$i = 1, \dots, n$$

# Estimation

**Minimize squared error (Least Squares Error)  
= Maximum Likelihood Estimation  
for linear model**

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

$$E(\hat{\beta}) = \beta$$

$$V(\hat{\beta}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

**Estimate  $\sigma^2$  by**

$$\hat{\sigma}^2 = \frac{\text{sum of squares error}}{n}$$

**or divide by  $n-1$   
for unbiased  
estimate**

# Inference – Model Comparison

Take linear model

$$\mathbf{y} = \mathbf{X}^T \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

And add constraint  $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$

this defines a new model that is a simplification of the previous one

# Inference – Model Comparison

**E.g., cf. model**  $y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + \varepsilon_i$

**to simplification with**  $\beta_3 = 0$

**i.e.**  $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$

$$(0,0,1) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 0$$

**i.e.**  $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$

**What about  $\beta_2 = 0$  &  $\beta_3 = 0$ ?**

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{c} \quad \Rightarrow \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

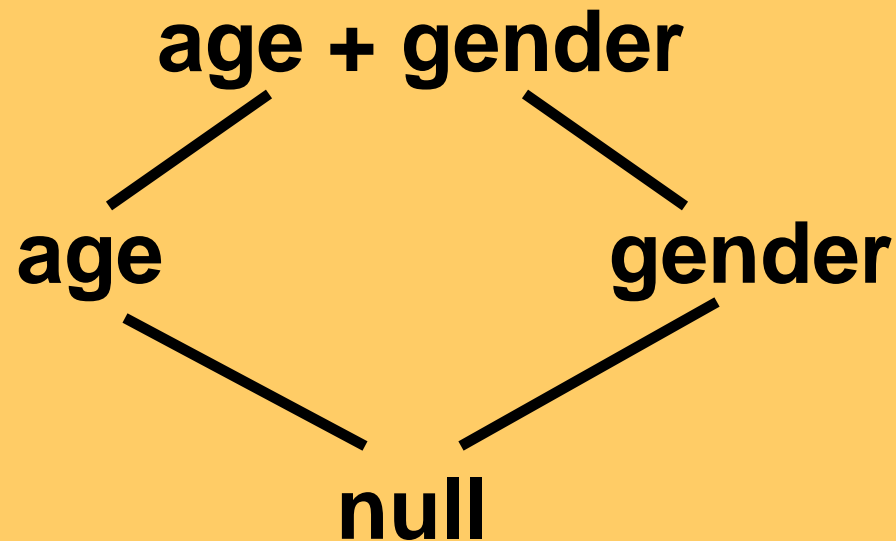
**And what about  $\beta_2 = \beta_3$ ?**

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{c} \quad \Rightarrow \quad \begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 0$$

**Are 2 different conditions equivalent?  
E.g. is the activation effect: reading a  
word vs imagining the object equal?**

**Definition: Linear model nested in another if 1<sup>st</sup> model can be obtained by linear constraint on the 2<sup>nd</sup>**

**Nesting tree:**



# F-test for General Linear Hypothesis

$$\mathbf{y} = \mathbf{X}^T \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim N_n \left( 0, \sigma^2 \mathbf{I}_n \right)$$

Consider

$$H_0 : \mathbf{A} \boldsymbol{\beta} = \mathbf{c}$$

**This is the General Linear Hypothesis**

Under  $H_0$  , i.e.,  $\mathbf{A}\boldsymbol{\beta} = \mathbf{c}$

$$F = \frac{(\text{Dev}_{\text{nested}} - \text{Dev}_{\text{larger}}) / (p_{\text{larger}} - p_{\text{nested}})}{(\text{Dev}_{\text{larger}}) / (n - p_{\text{larger}})} : F_{p_{\text{larger}} - p_{\text{nested}}, n - p_{\text{larger}}}$$

$p$  denotes the number of model parameters

$n$  denotes the number of data points

$Dev$  = Deviance = sum of squares of residuals

Tests ratio of variances

# FMRI Data:

**Set of Volumes (over time) or  
Set of Time-Series (over space)**

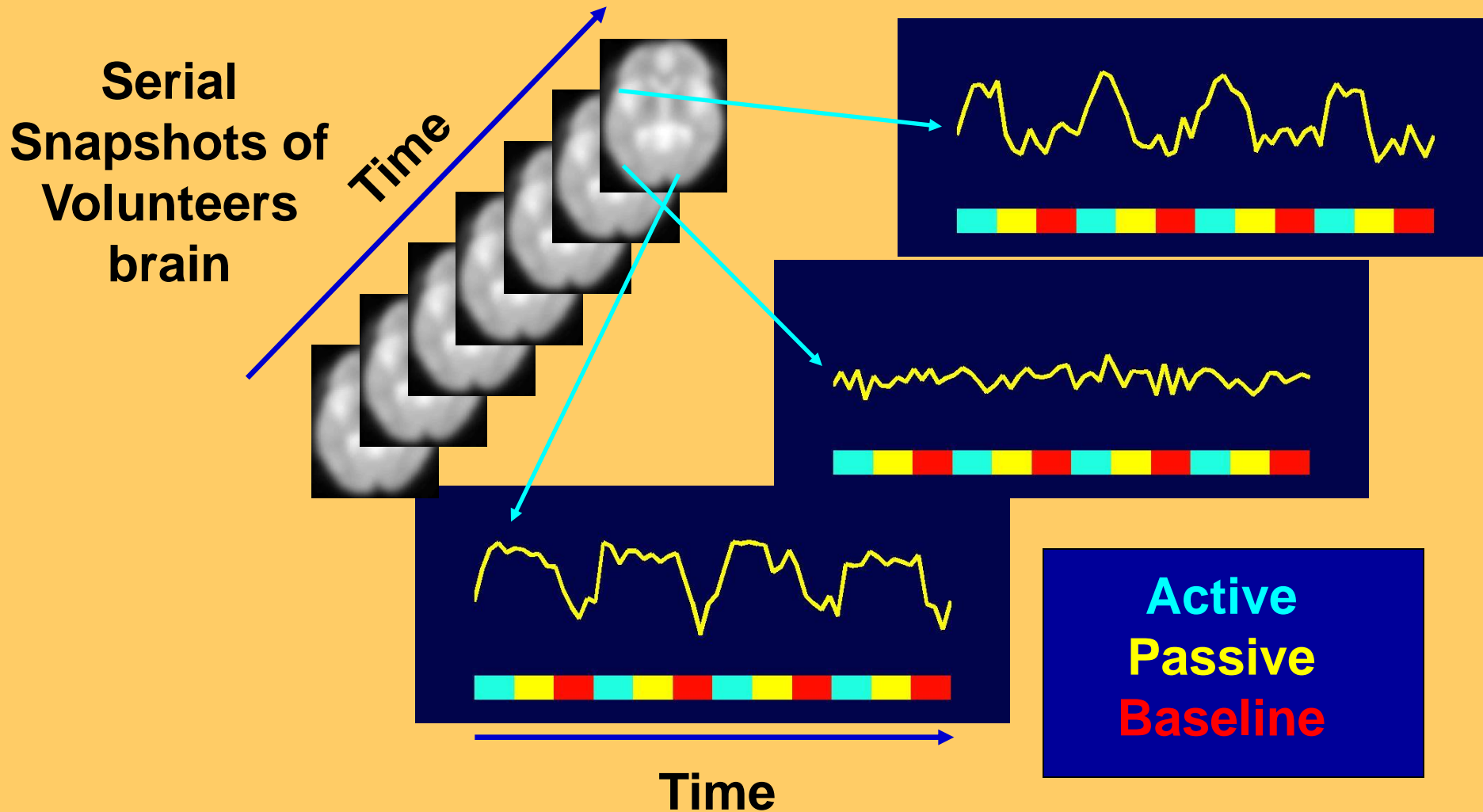
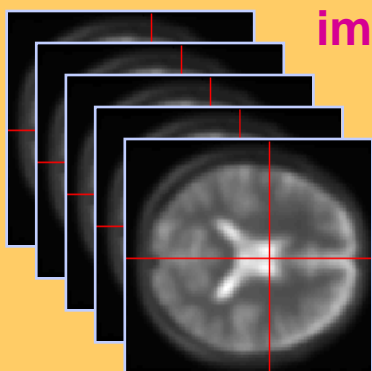
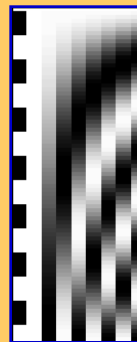


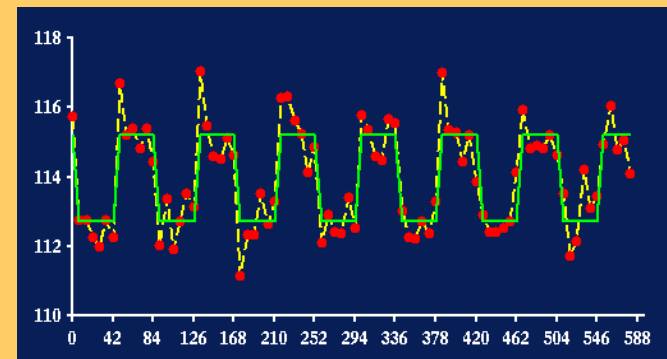
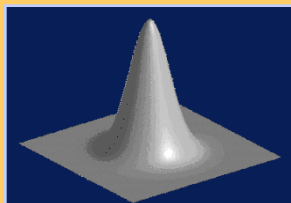
image data



design matrix



kernel



parameter estimates

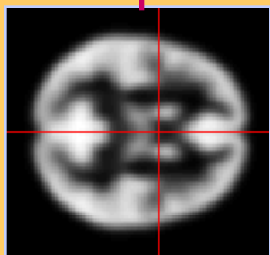
realignment & motion correction

smoothing

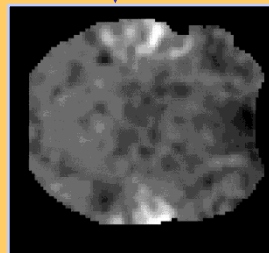
Linear Model  
 ↳ model fitting  
 ↳ statistic image

Thresholding & Random Field Theory

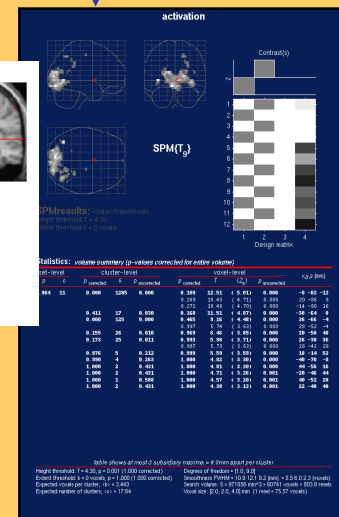
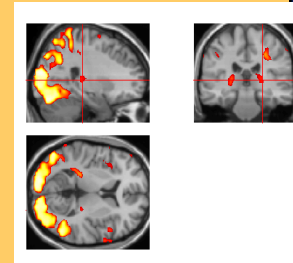
normalisation



anatomical reference



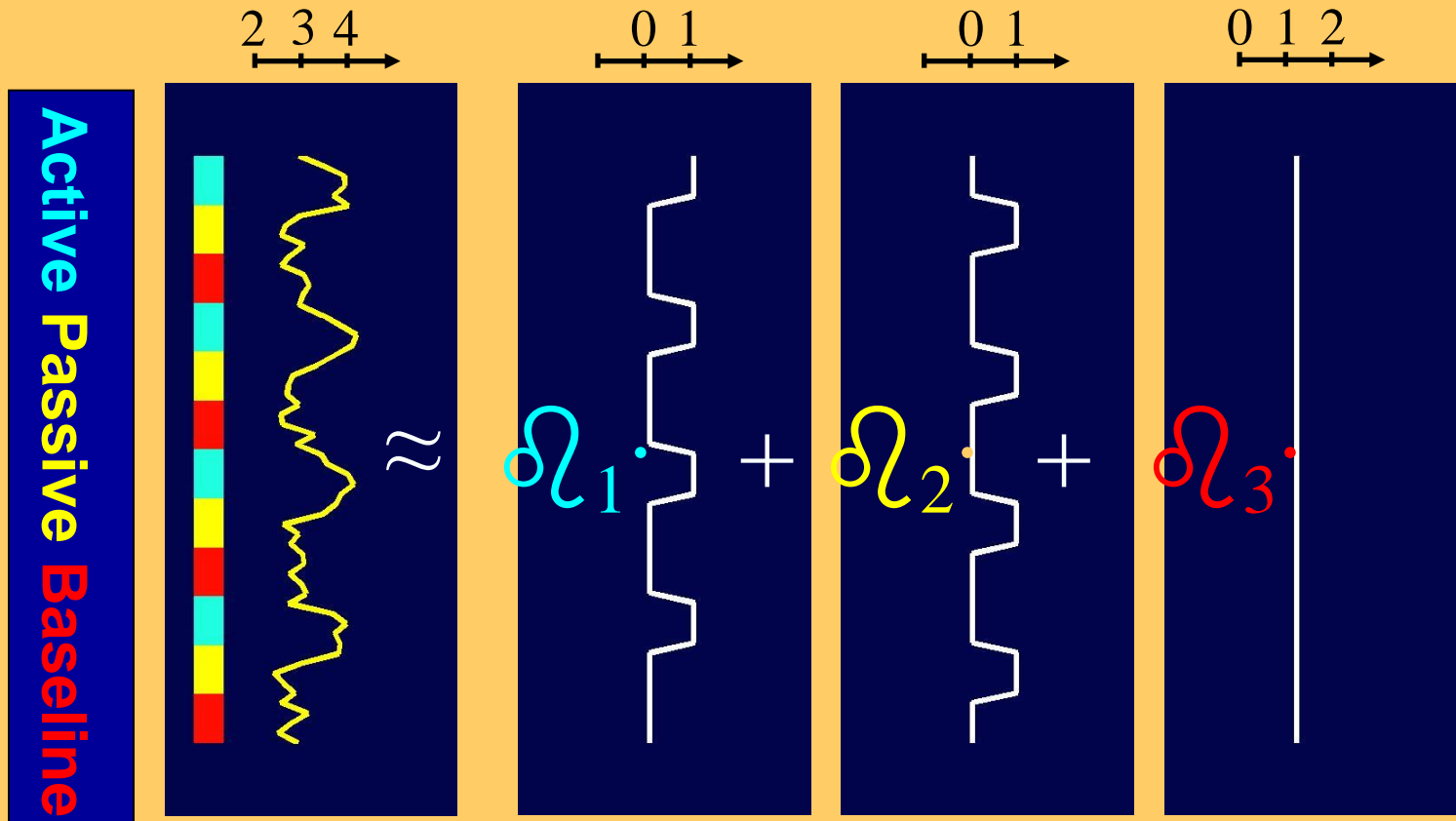
Statistical Parametric Map (test statistics)



Corrected thresholds & p-values

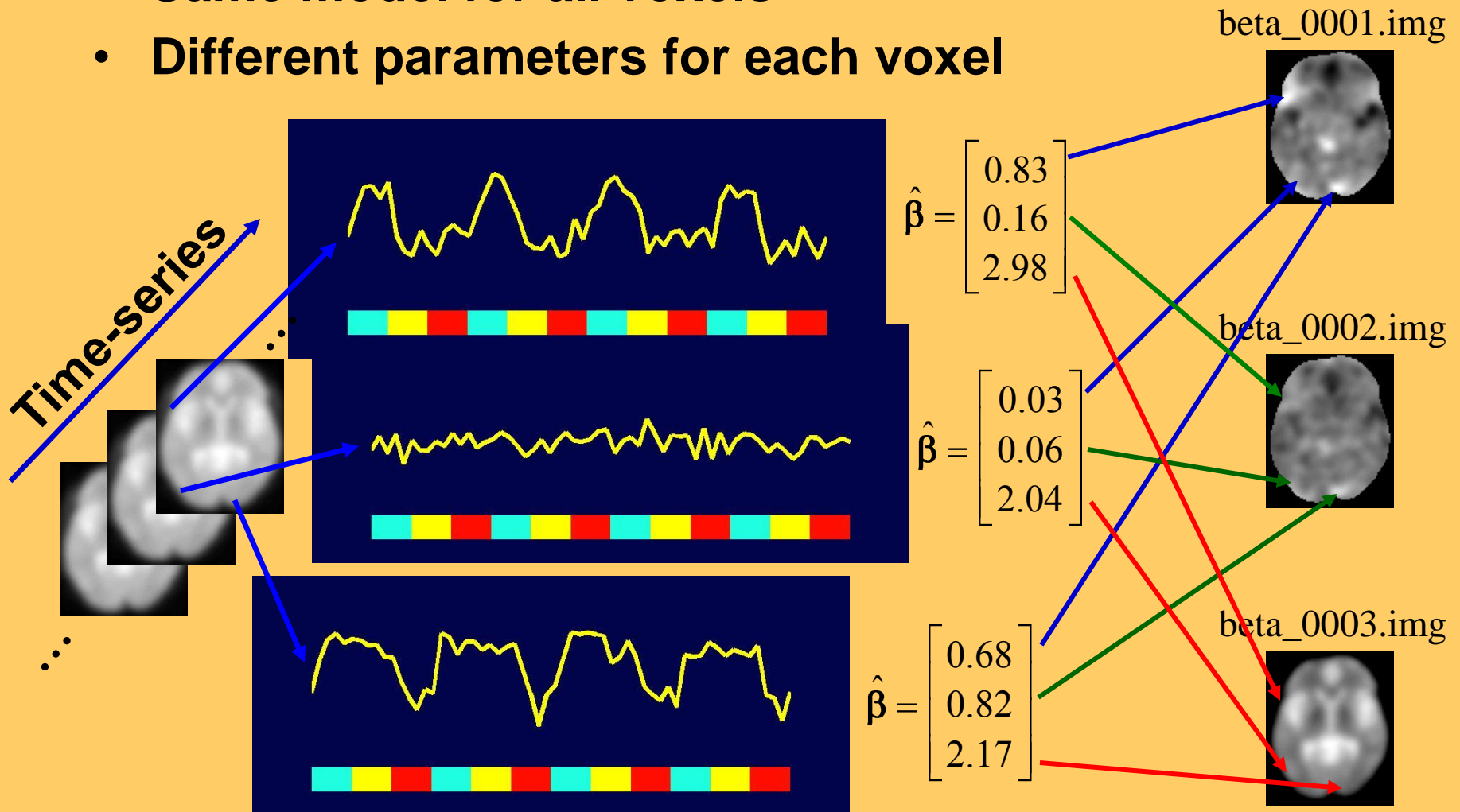
# Estimation

The estimation entails finding the parameter values such that the linear combination *best* fits the data



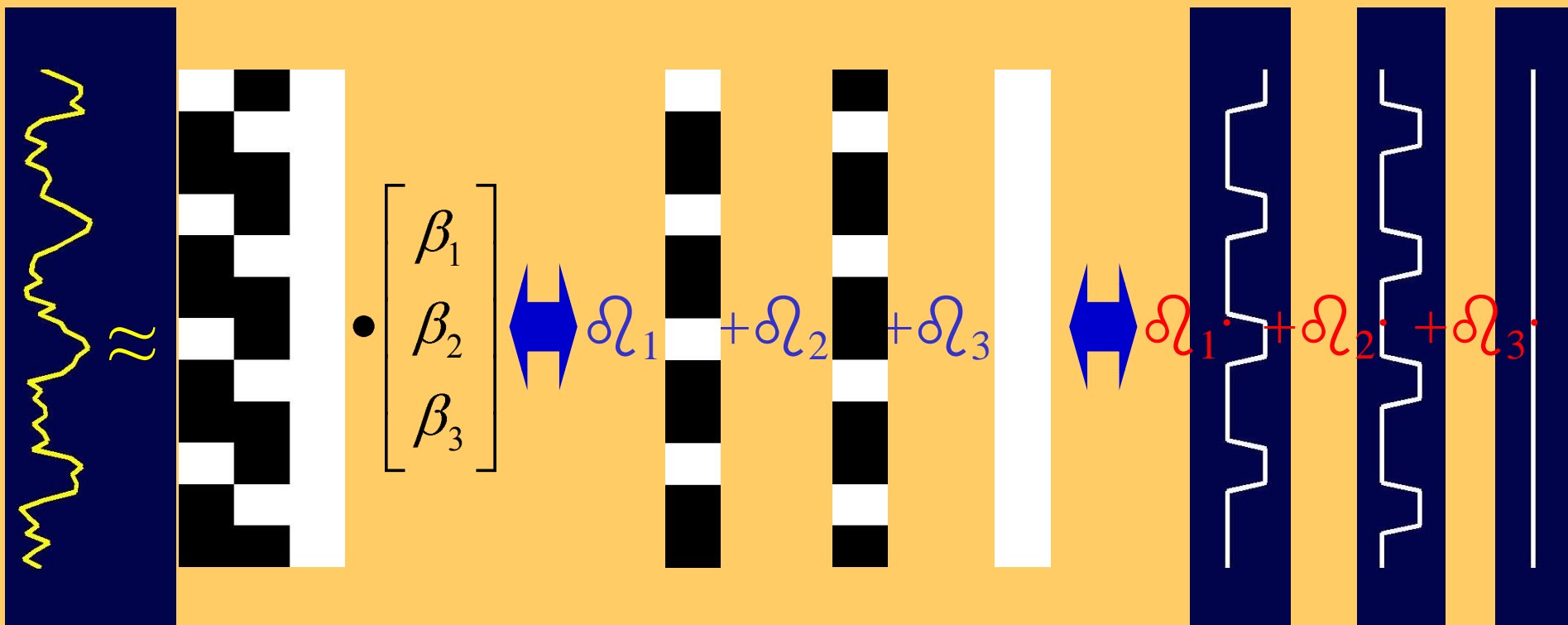
# Parameter Estimates

- Same model for all voxels
- Different parameters for each voxel



$$\mathbf{y} \approx \mathbf{X}^T \boldsymbol{\beta}$$

**SPM View**



**Note:**

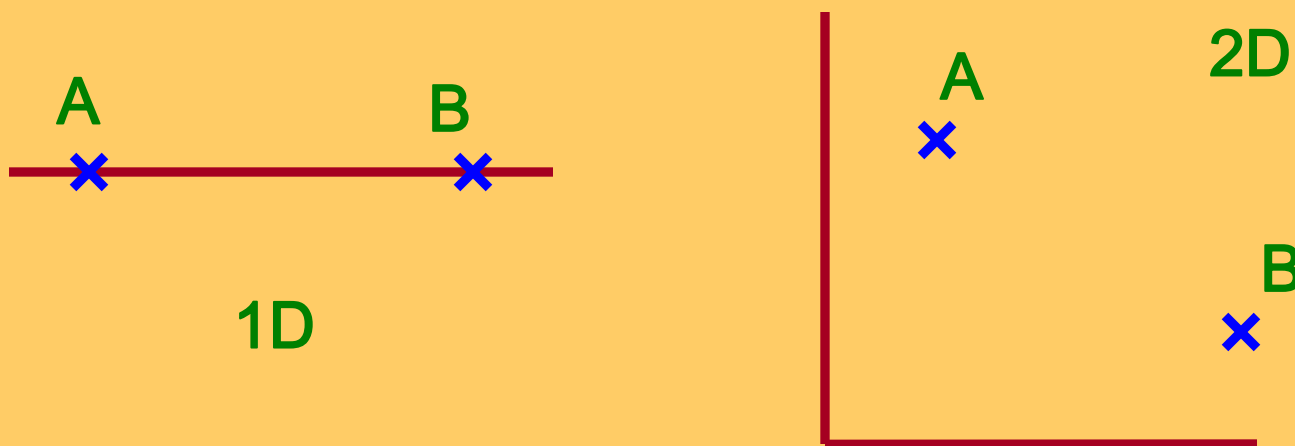
**We trust: Long series with large effects and small error**

# Spatial Modeling

# Spatial Hypotheses

**Question - how do we extend from standard univariate hypotheses to answering spatially motivated questions?**

**Not easy - curse of dimensionality (millions of voxels)**



**in 1D it makes sense to infer A is less than B, but what is the equivalent in 2D?**

# Spatial Testing Solutions

- Summarize the image into one dimensional quantities for testing (e.g. *region of interest analysis*)
- Consider the overall test as a combination of individual voxel tests (*voxel based analysis*)
- Perform *shape/object analysis* on objects defined via landmarks
- Build *Bayesian image analysis* models

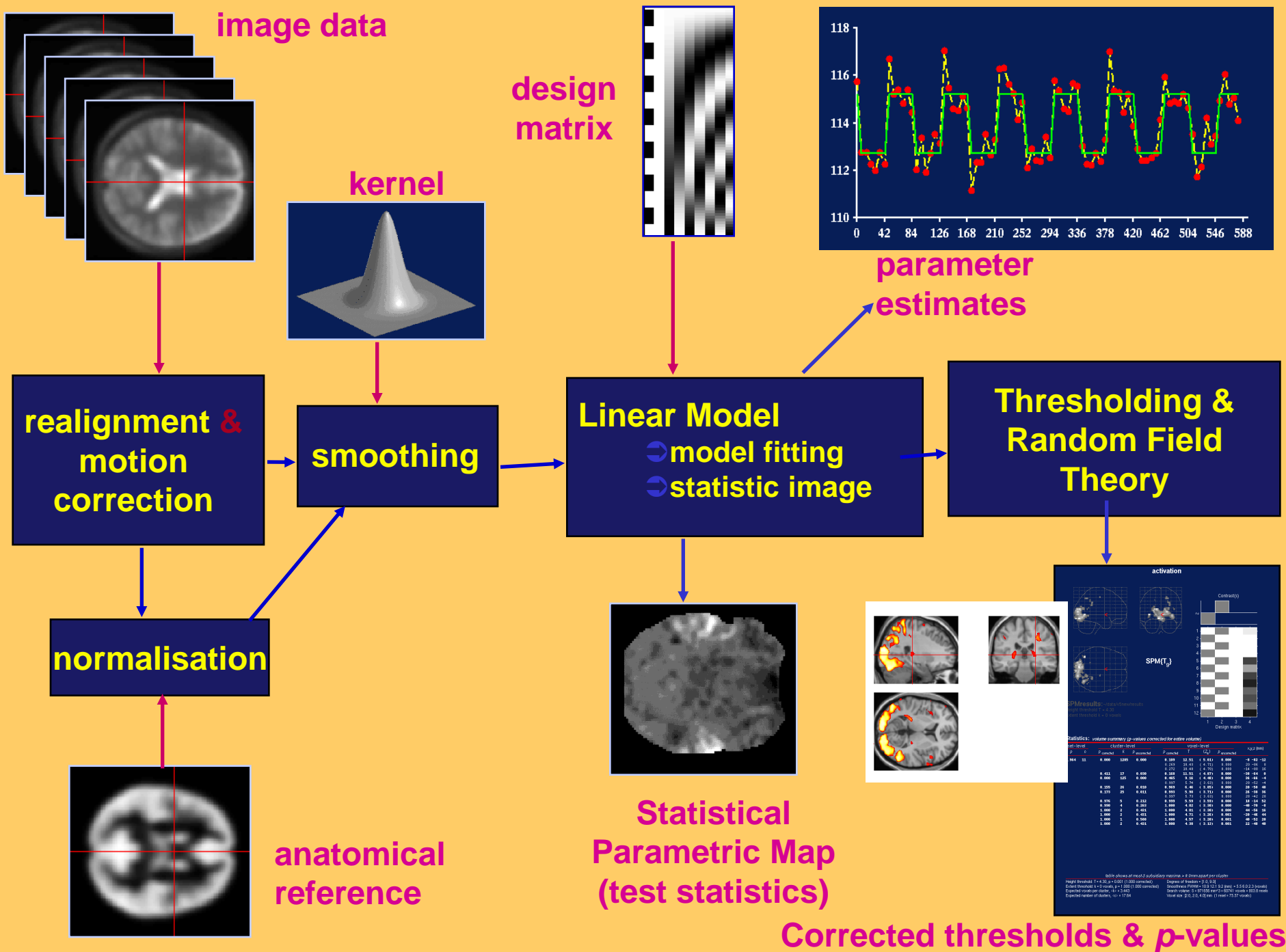
# Spatial Testing Solutions

- Summarize the image into one dimensional quantities for testing (e.g. *region of interest analysis*)
- Consider the overall test as a combination of individual voxel tests (*voxel based analysis*)
- Perform *shape/object analysis* on objects defined via landmarks
- Build *Bayesian image analysis* models

# Voxel based analysis

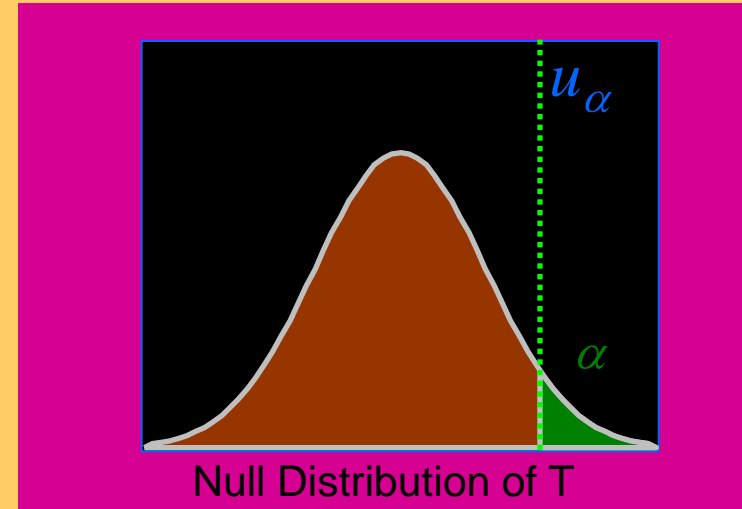
Each voxel obtains a test statistic from the linear model, e.g.  $t$  or  $F$

Forms statistical maps of the statistics



# Hypothesis Testing

- Null Hypothesis  $H_0$
- Test statistic  $T$ 
  - $t$  observed realization of  $T$
- $\alpha$ -level
  - Acceptable false positive risk
  - Level  $\alpha = \Pr( T > u_\alpha \mid H_0 )$
  - Threshold  $u_\alpha$  controls false positive risk at level  $\alpha$



# Multiple Comparisons Problem

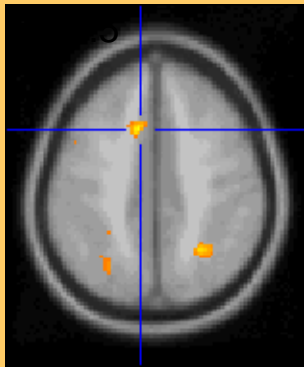
**Which of 100,000 voxels are significant?**

–  $\alpha = 0.05 \Rightarrow$  **5,000 false positive voxels**

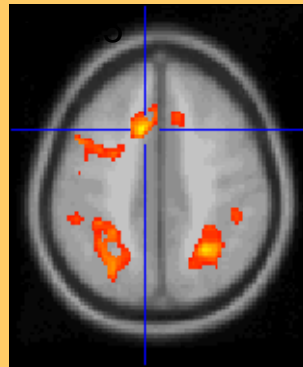
# Assessing Statistic Images

Where's the signal or change?

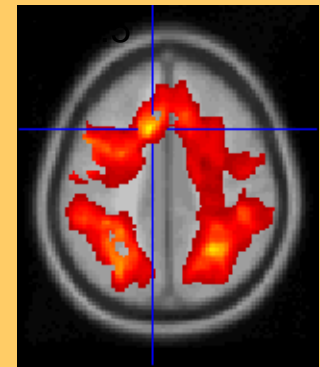
**High Threshold**



**Med. Threshold**



**Low Threshold**



**Good Specificity**

**Poor Power**  
(risk of false  
negatives)

**Poor Specificity**  
(risk of false  
positives)

**Good Power**

**How can we determine a sensible threshold level?**

# Multiple Comparison Solutions: Measuring False Positives

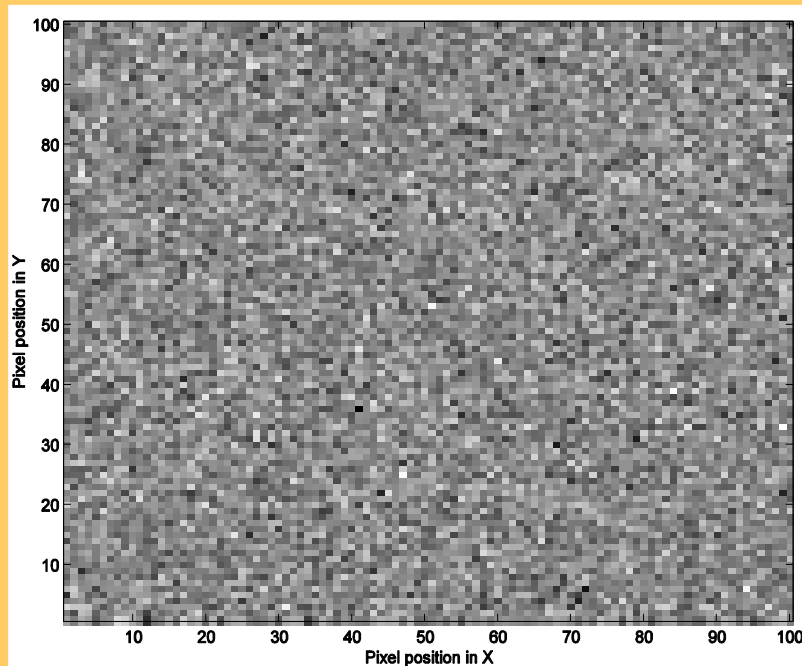
- **Familywise Error Rate (FWER)**
  - **Familywise Error**
    - Existence of one or more false positives
- **False Discovery Rate (FDR)**
  - **$FDR = E(V/R)$**
  - **R voxels declared active, V falsely so**  
**Realized false discovery rate:  $V/R$**

# Bonferroni Correction

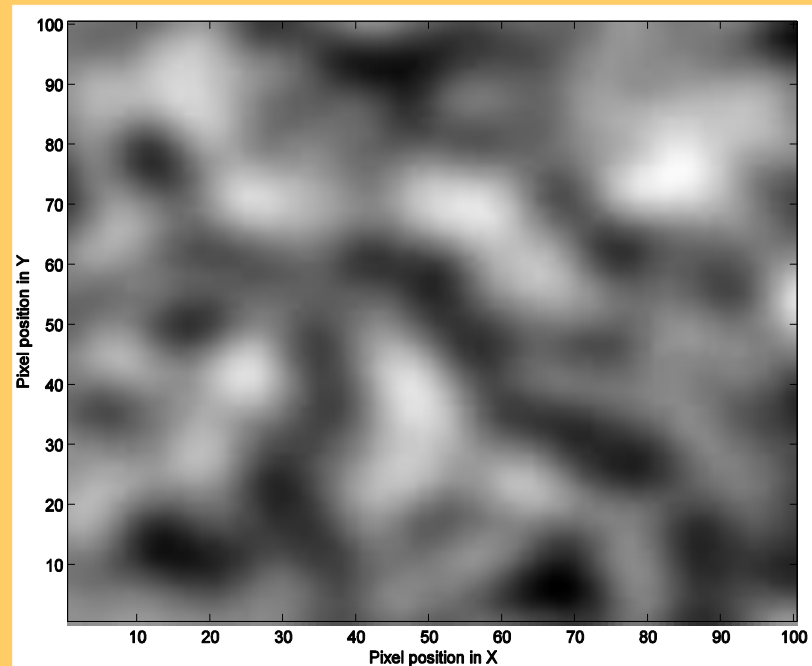
FWE,  $\alpha$ , for  $N$  **independent** voxels is  $\alpha = Nv$  ( $v$  = voxel-wise error rate)

To control FWE set  $v = \alpha / N$

Independent Voxels



Spatially Correlated Voxels



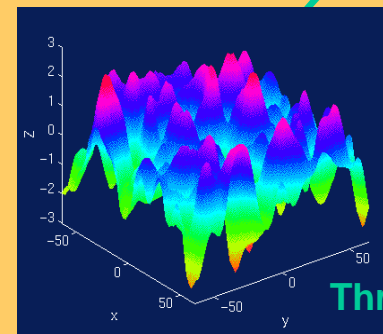
Bonferroni is too conservative for brain images

# FWER MCP Solutions: Random Field Theory

- **Euler Characteristic  $\chi_u$** 
  - Topological Measure
    - #blobs - #holes
  - At high thresholds, just counts blobs
  - **FWER** =  $\Pr(\text{Max voxel} \geq u \mid H_o)$   
 =  $\Pr(\text{One or more blobs} \mid H_o)$   
 $\approx \Pr(\chi_u \geq 1 \mid H_o)$   
 $\approx E(\chi_u \mid H_o)$

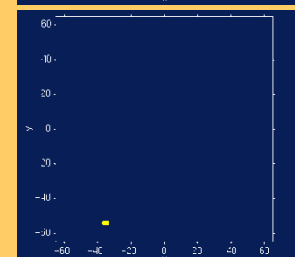
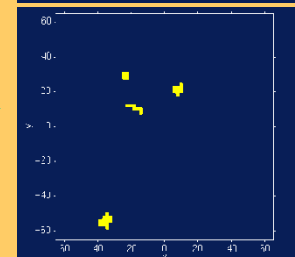
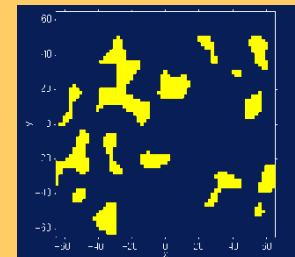
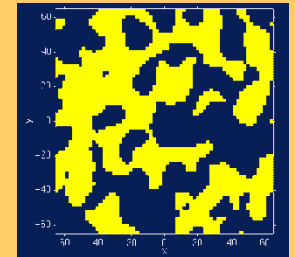
*No holes*

*Never  
more than  
1 blob*



Random Field

Threshold

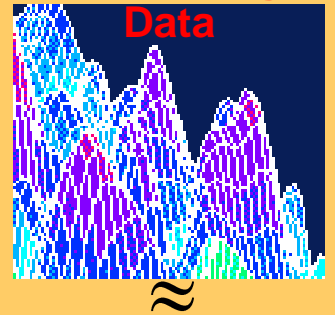


# Random Field Theory

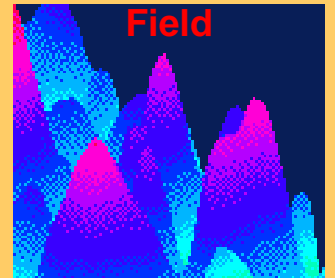
## Limitations

- **Multivariate normality (Gaussianity)**
  - Virtually impossible to check
- **Sufficient smoothness**
  - FWHM smoothness 3-4 voxel size
- **Smoothness estimation**
  - Estimate is biased when images not sufficiently smooth
- **Several layers of approximations**

Lattice Image  
Data



Continuous Random  
Field



# Multiple Comparisons Solutions: Measuring False Positives

- **Familywise Error Rate (FWER)**
  - Familywise Error
    - Existence of one or more false positives
  - FWER is probability of familywise error
- **False Discovery Rate (FDR)**
  - $FDR = E(V/R)$
  - $R$  voxels declared active,  $V$  falsely so
    - Realized false discovery rate:  $V/R$

# False Discovery Rate

- For any threshold, all voxels can be cross-classified:

	Accept Null	Reject Null
Null True	$V_{0A}$	$V_{0R}$
Null False	$V_{1A}$	$V_{1R}$
	$N_A$	$N_R$

- Realized FDR

$$\text{rFDR} = V_{0R} / (V_{1R} + V_{0R}) = V_{0R} / N_R$$

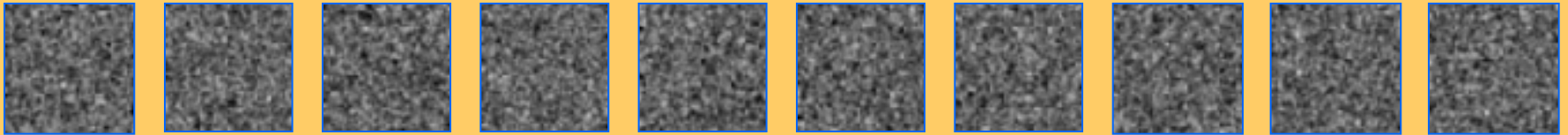
– If  $N_R = 0$ ,  $\text{rFDR} = 0$

- But only can observe  $N_R$ , don't know  $V_{1R}$  &  $V_{0R}$ 
  - We control the *expected* rFDR

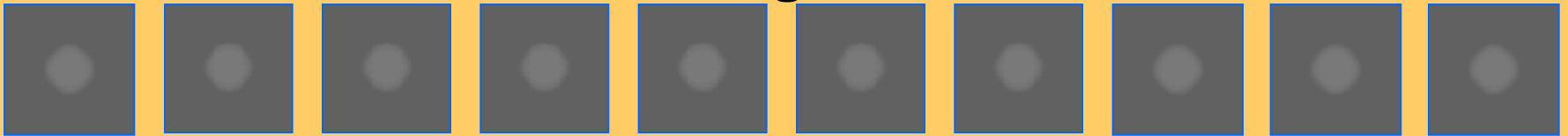
$$\text{FDR} = \text{E}(\text{rFDR})$$

# False Discovery Rate Illustration:

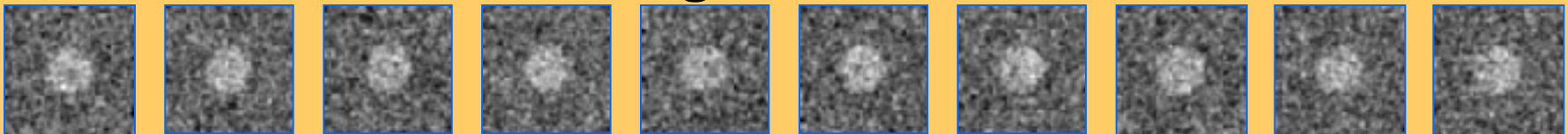
Noise



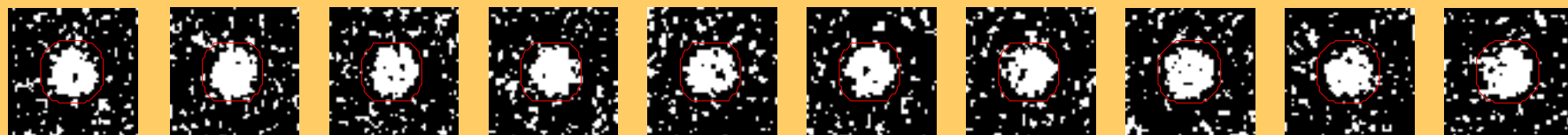
Signal



Signal+Noise



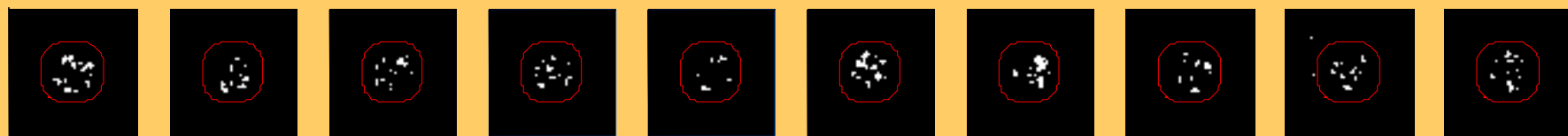
## Control of Per Comparison Rate at 10%



11.3% 11.3% 12.5% 10.8% 11.5% 10.0% 10.7% 11.2% 10.2% 9.5%

Percentage of Null Pixels that are False Positives

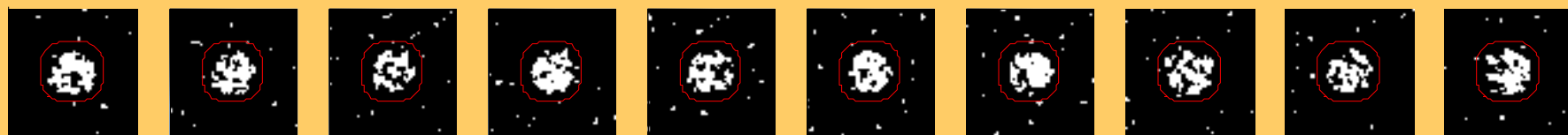
## Control of Familywise Error Rate at 10%



FWE

Occurrence of Familywise Error

## Control of False Discovery Rate at 10%



6.7% 10.4% 14.9% 9.3% 16.2% 13.8% 14.0% 10.5% 12.2% 8.7%

Percentage of Observed "Above Threshold" Pixels that are False Positives

# Benjamini & Hochberg Procedure

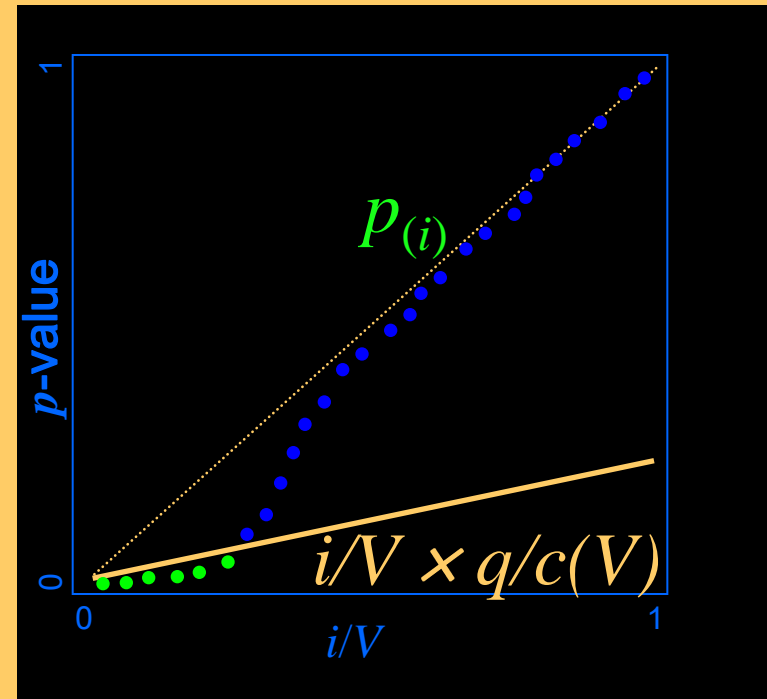
*Journal of the Royal  
Statistical Society – Series B*  
(1995) 57:289-300

- Select desired limit  $q$  on FDR
- Order p-values,  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(V)}$
- Let  $r$  be largest  $i$  such that

$$p_{(i)} \leq i/V \times q/c(V)$$

- Reject all hypotheses corresponding to

$$p_{(1)}, \dots, p_{(r)}$$



**NB, no spatial consideration**

# Also, Non-Parametric Testing

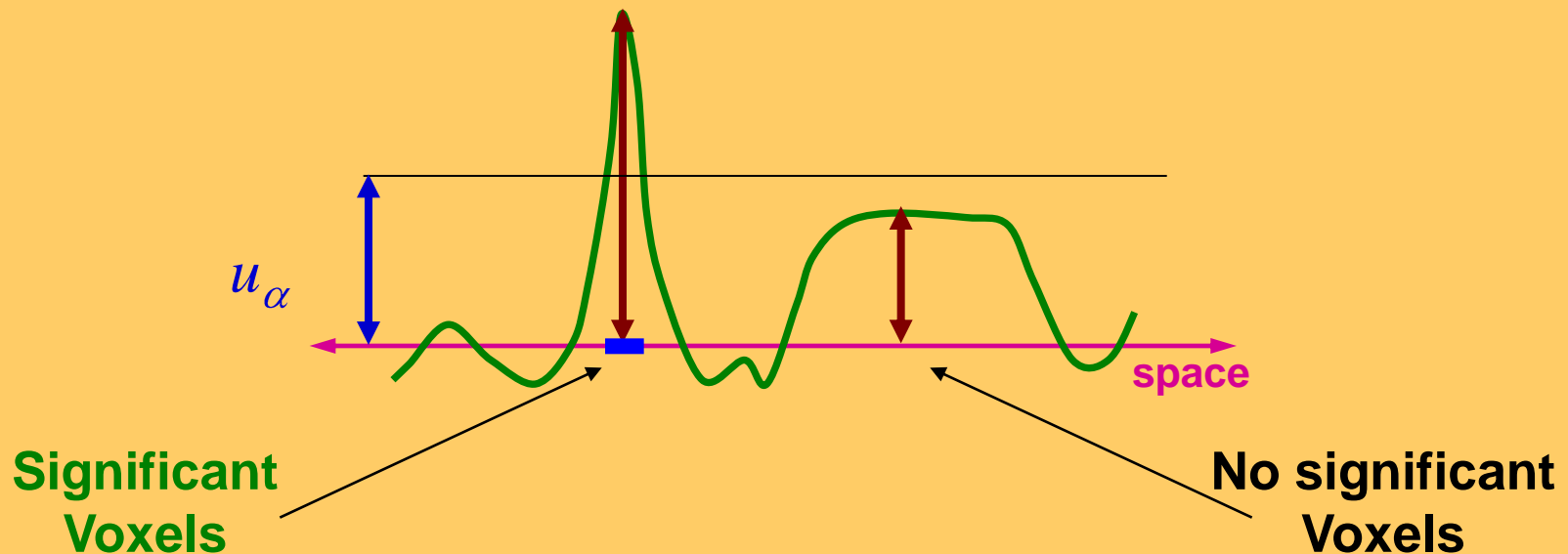
- If  $H_0$  is true then time order irrelevant (if noise really iid)
- Therefore permute the timepoints and obtain test statistics
- If true test statistic is extreme compared to others then reject  $H_0$

# Types of Spatial Inference

- **Individual voxel level**
- **Cluster level**
- **Set level**
- **Bayesian model based**

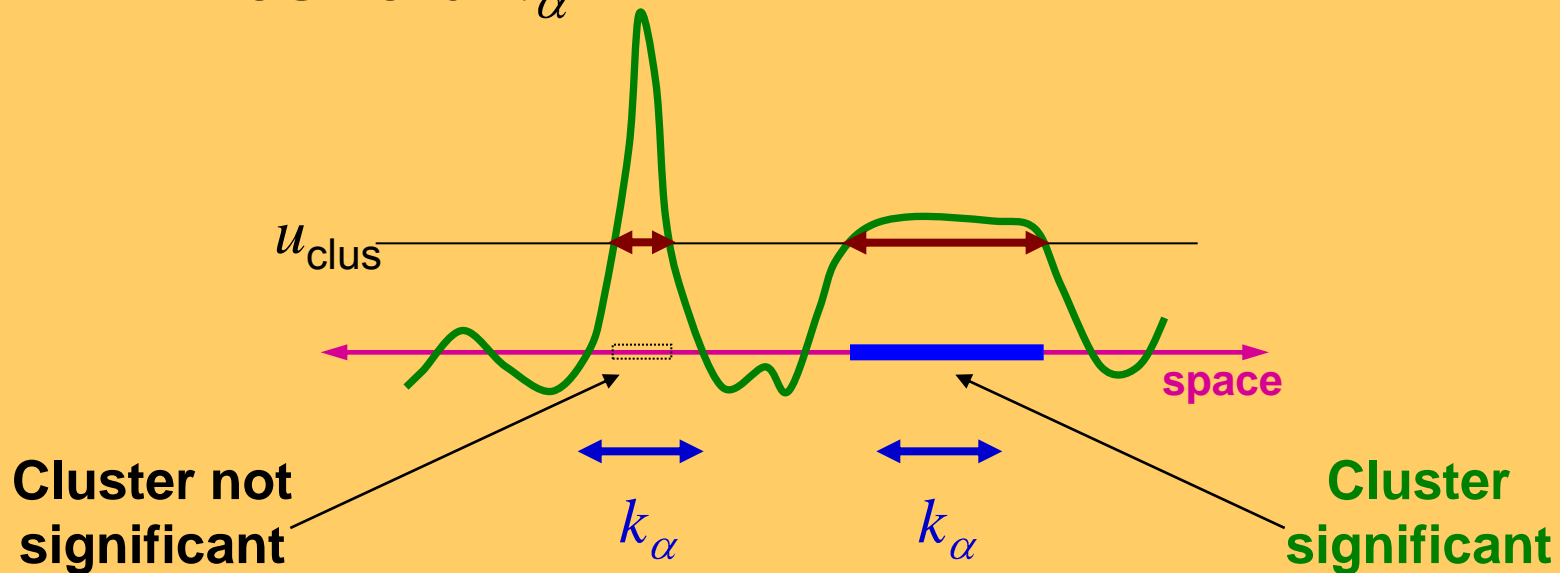
# Voxel-level Inference

- Retain voxels above  $\alpha$ -level threshold  $u_\alpha$
- Gives best spatial specificity
  - $H_0$  at a single voxel can be rejected



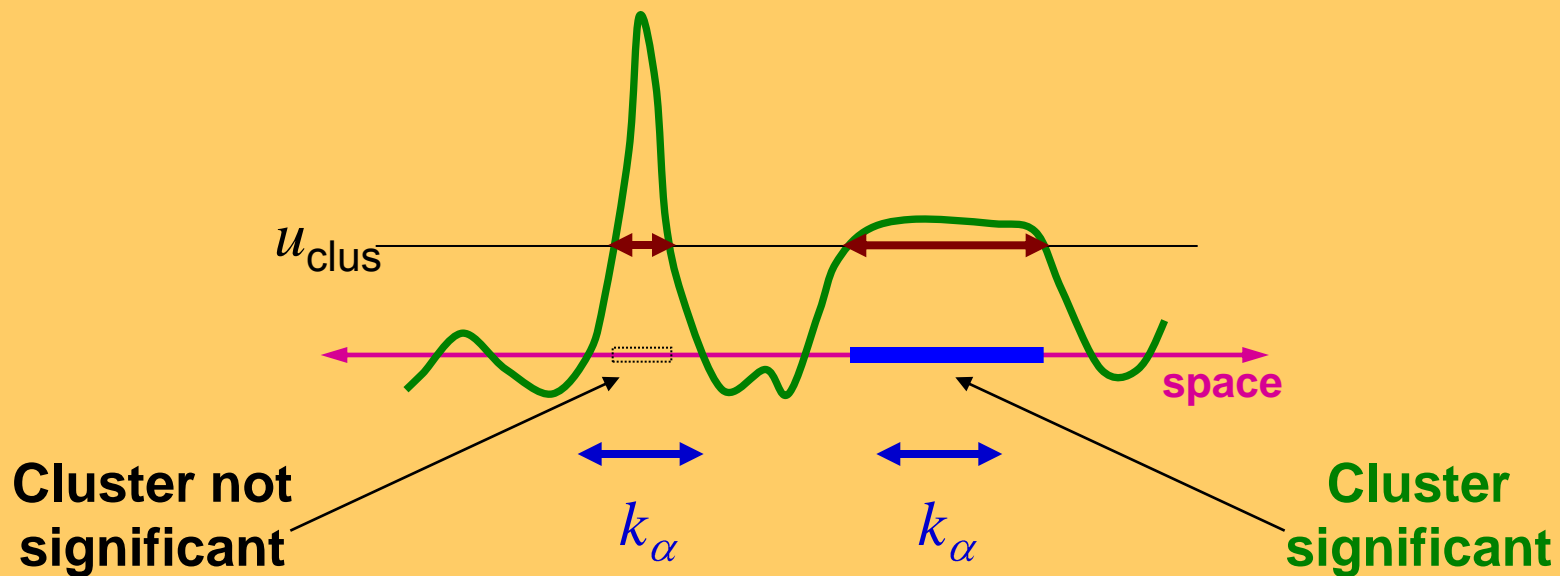
# Cluster-level Inference

- **Two step-process**
  - Define clusters by arbitrary threshold  $u_{\text{clus}}$
  - Retain clusters larger than  $\alpha$ -level threshold  $k_\alpha$



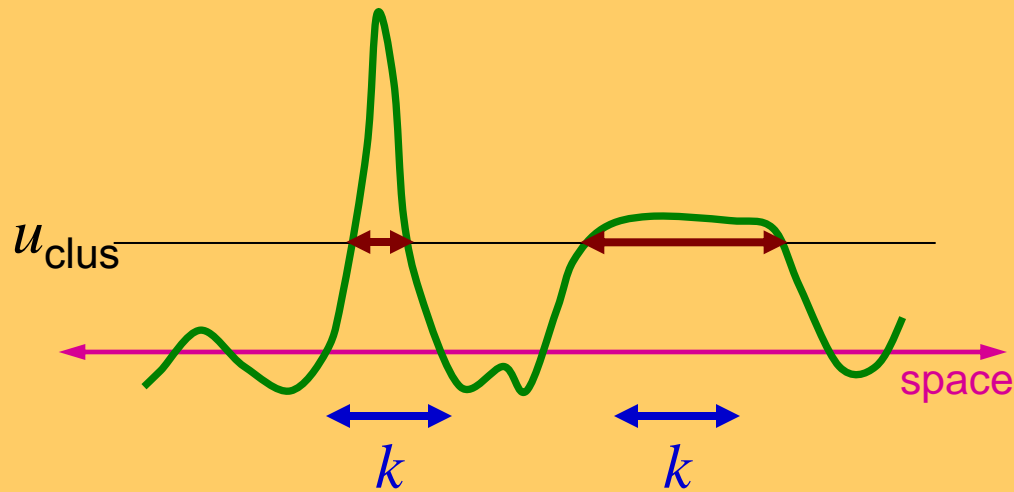
# Cluster-level Inference

- Typically better sensitivity
- Worse spatial specificity
  - The null hyp. of entire cluster is rejected
  - Only means that *one or more* of voxels in cluster active



# Set-level Inference

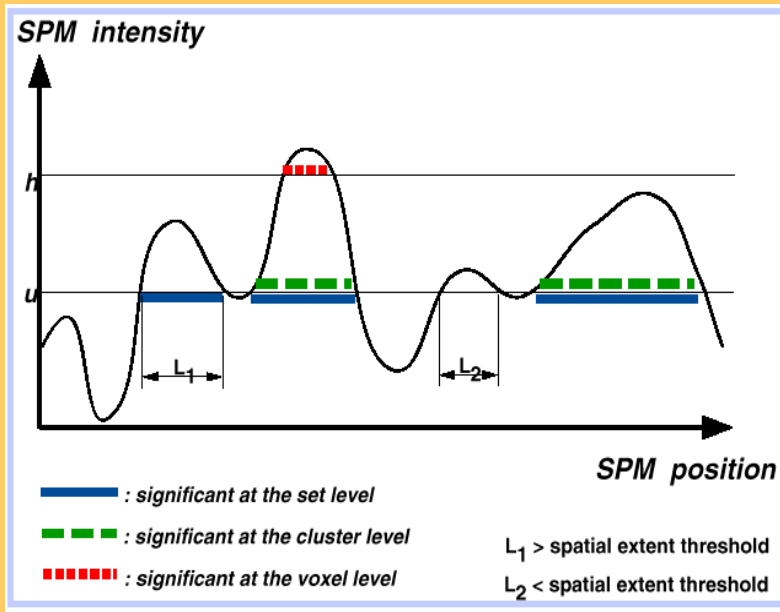
- Count number of blobs  $c$ 
  - Minimum blob size  $k$
- Worst spatial specificity
  - Only can reject global null hypothesis



Here  $c = 1$ ; only 1 cluster larger than  $k$

# Review:

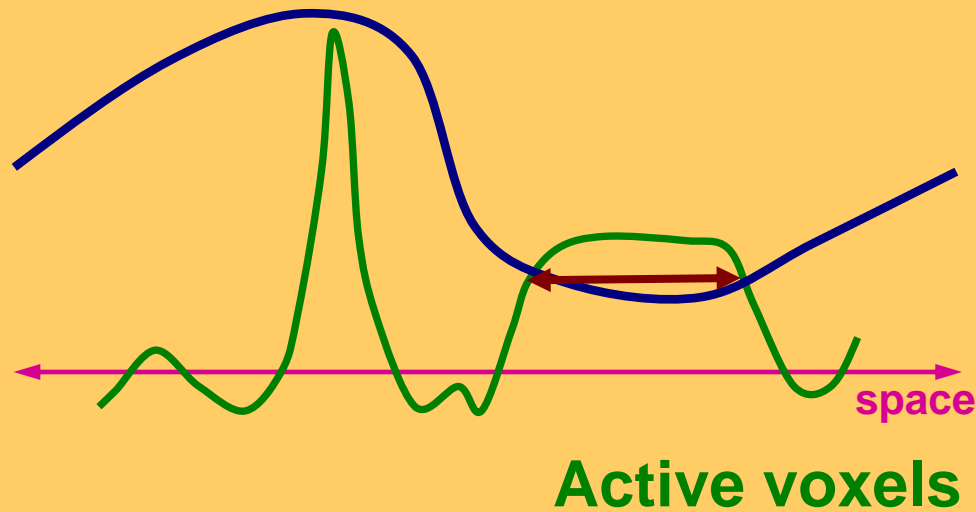
## Levels of inference & power



Sensitivity	Test based on	Parameters set by the user	Regional specificity
⊖	The intensity of a voxel	• Low pass filter	⊕
	The spatial extent above $u$ or the spatial extent and the maximum peak height	• Low pass filter • intensity threshold $u$	
	The number of clusters above $u$ with size greater than $n$	• Low pass filter • intensity thres. $u$ • spatial threshold $n$	
⊕	The sum of square of the SPM or a MANOVA	• Low pass filter	⊖

# A flexible Bayesian Approach

- Model the form of activity
- Provides an “adaptive thresholding” approach



# Bayesian Model

$$y = zx + \varepsilon$$

$y$  = data, parameter estimates of statistics

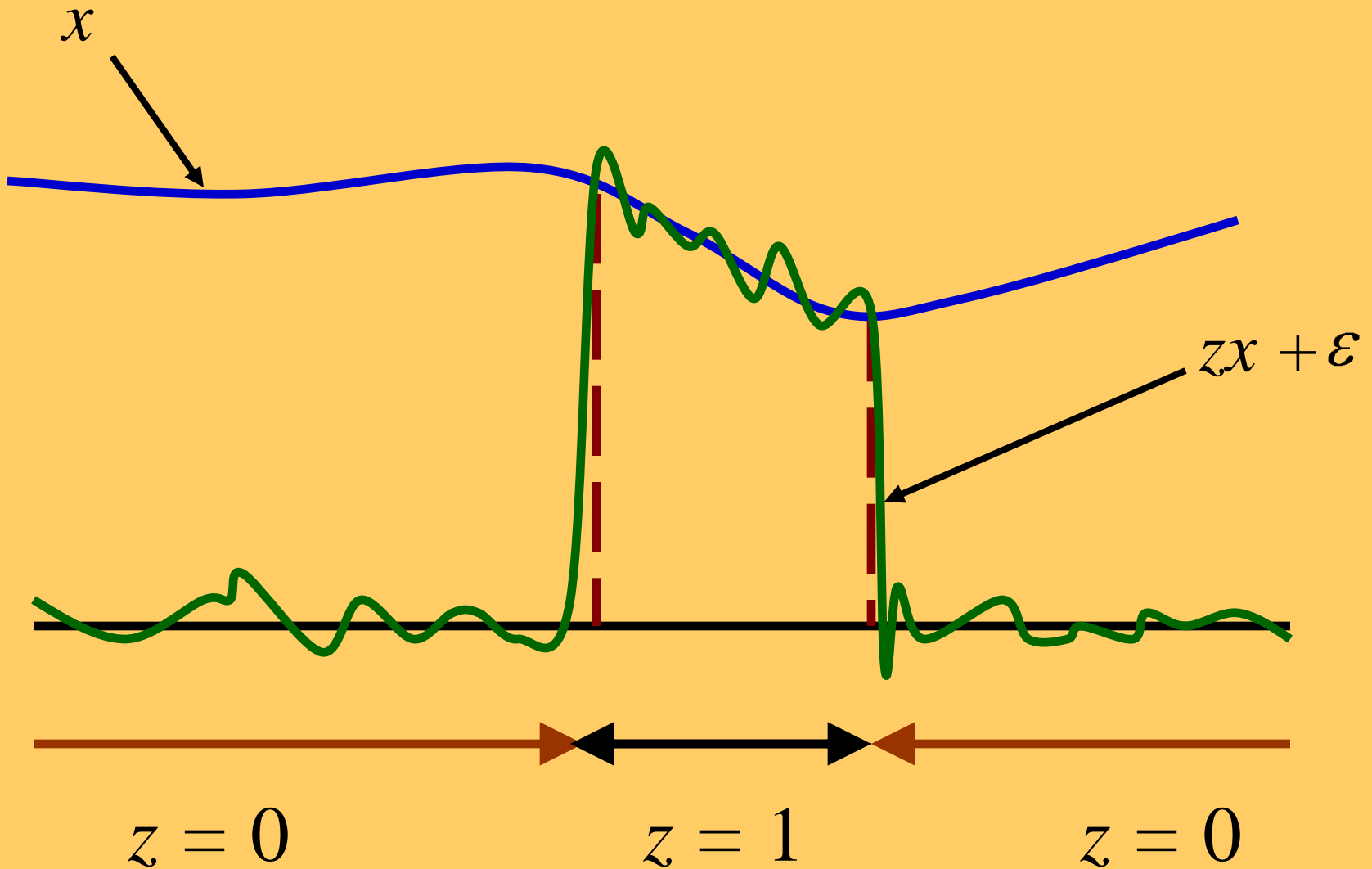
$z$  = binary activation map – modeled as a MRF

$x$  = activation level field – modeled as a MRF

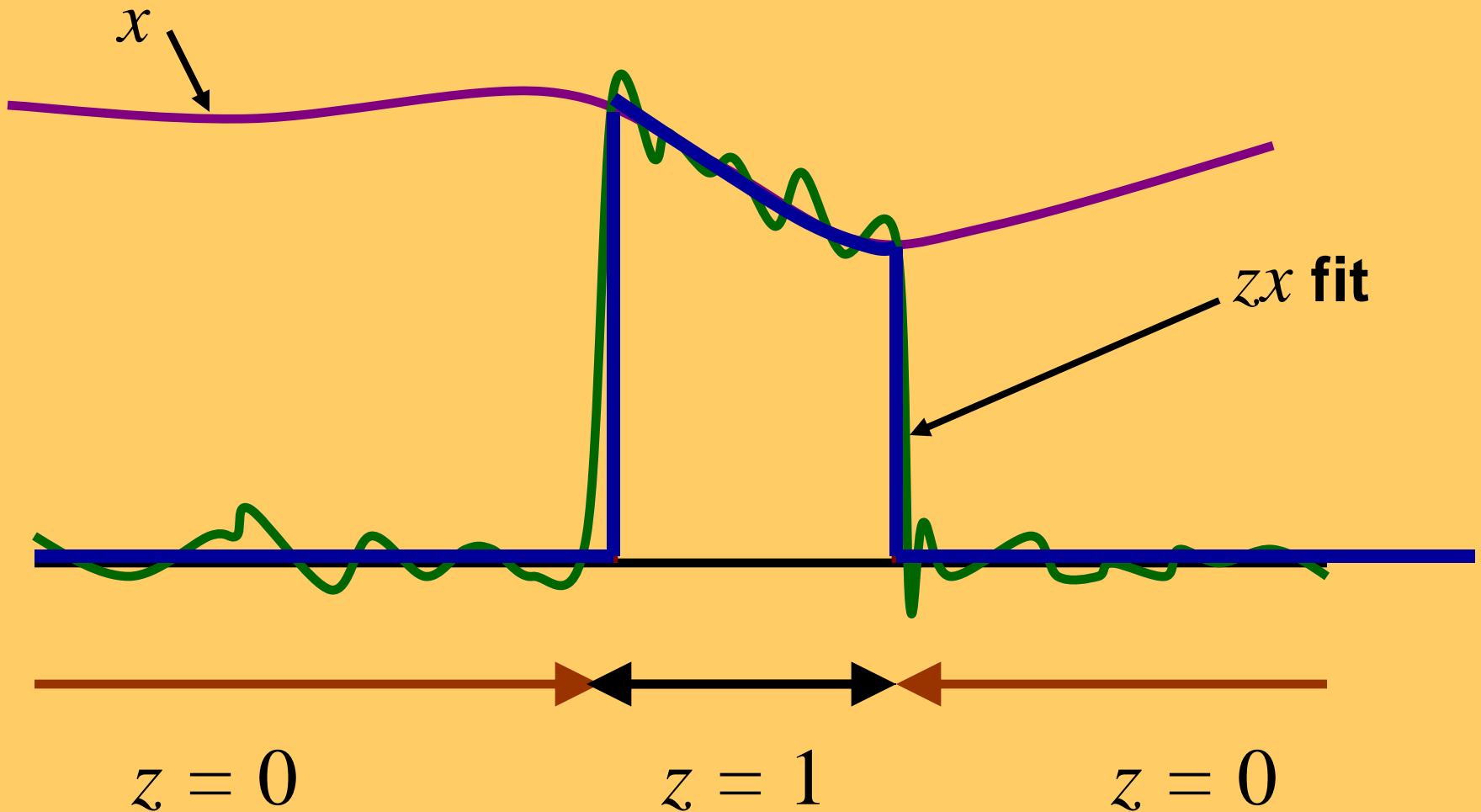
$\varepsilon$  = residual error

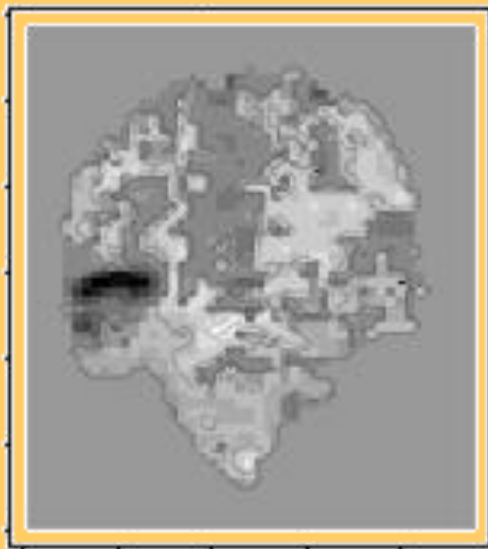
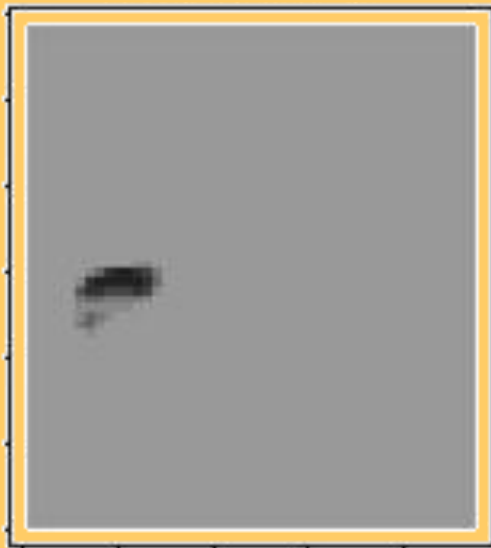
**MRF = Markov Random Field (similar random field but defined on a lattice)**

# Model Illustration



# Model Illustration



$y$  $x$  $z$  $zx$  $+$  $\varepsilon$  $=$ 

# Other Topics and Omissions

- Hemodynamic response function
- Multiple subjects (random and mixed effects models)
- PCA, ICA
- Multivariate analysis with variogram modeling
- Space-time modeling

# Plug

**Spring lecture series:**

**“Statistics for Radiology and Biomedical Imaging” - China Basin Landing Classroom**  
**- Spring Quarter**

**This will probably run again in 2011**